MNRAS **000**, 1–20 (2015) Preprint 14 October 2015 Compiled using MNRAS LaTeX style file v3.0

Comparing models of the periodic variations in spin-down and beam-width for PSR B1828-11

G. Ashton,[1⋆] D.I. Jones,[1] R. Prix[2]
[1] *Mathematical Sciences and STAG Research Centre, University of Southampton, Southampton SO17 1BJ*
[2] *Max Planck Institut für Gravitationsphysik (Albert Einstein Institut), 30161 Hannover, Germany*

14 October 2015

ABSTRACT

We build a framework using tools from Bayesian data analysis to evaluate models explaining the periodic variations in spin-down and beam-width of PSR B1828-11. The available data consists of the time averaged spin-down rate, which displays a distinctive double-peaked modulation, and measurements of the beam-width. Two concepts exist in the literature that are capable of explaining these variations; we will formulate predictive models from these and quantitatively compare them. The first concept is phenomenological and stipulates that the magnetosphere undergoes periodic switching between two meta-stable states as first suggested by Lyne et al. (2010). The second concept, precession, was first considered as a candidate for the modulation of B1828-11 by Stairs et al. (2000). We quantitatively compare models built from these concepts using a Bayesian odds-ratio. Because the phenomenological switching model itself was informed by this data in the first place, it is difficult to specify appropriate parameter-space priors that can be trusted for an unbiased model comparison. Therefore we first perform a parameter estimation using the spin-down data, and then use the resulting posterior distributions as priors for model comparison on the beam-width data. We find that a precession model with a simple circular Gaussian beam geometry fails to appropriately describe the data, while allowing for a more general beam geometry results in a model that seems strongly preferred by the data over a switching model.

Key words: methods: data analysis – pulsars: individual: PSR B1828-11 – stars: neutron

1 INTRODUCTION

The pulsar B1828-11 demonstrates periodic variability in its pulse timing and beam shape at harmonically related periods of 250, 500, and 1000 days. The modulations in the timing was first taken as evidence that the pulsar is orbited by a system of planets by Bailes et al. (1993). A more complete analysis by Stairs et al. (2000) concluded that the corresponding changes in the beam-shape would require at least two of the planets to interact with the magnetosphere, which does not seem credible. Instead the authors proposed that the correlation between timing data and beam-shape suggested the pulsar was undergoing free precession. If true, such a claim would require rethinking of the vortex-pinning model used to explain the pulsar glitches since the pinning should lead to much shorter modulation period than observed (Shaham 1977), and fast damping of the modulation (Link 2003).

The idea of precession for B1828-11 has been studied extensively in the literature: Jones & Andersson (2001) derived the observable modulations due to precession and noted that the electromagnetic spin-down torque will amplify these modulations. Link & Epstein (2001) fitted a torqued-precession model to the spin-down and beam-shape followed by Akgün et al. (2006) where a variety of shapes and the form of the spin-down torque were tested. All of these authors agree that precession is a credible candidate to explain the observed periodic variations: furthermore to explain the double-peaked spin-down modulations, the so-called wobble angle must be small while the magnetic dipole must be close to $\pi/2$.

More recently Arzamasskiy et al. (2015) updated the previous estimates (based on a vacuum approximation) to a plasma filled magnetosphere. They also find that the magnetic dipole and spin-vector must be close to orthogonal, but solutions could exist where it is the wobble angle which is close to $\pi/2$ while the magnetic dipole lies close to the an-

⋆ E-mail: G.Ashton@soton.ac.uk

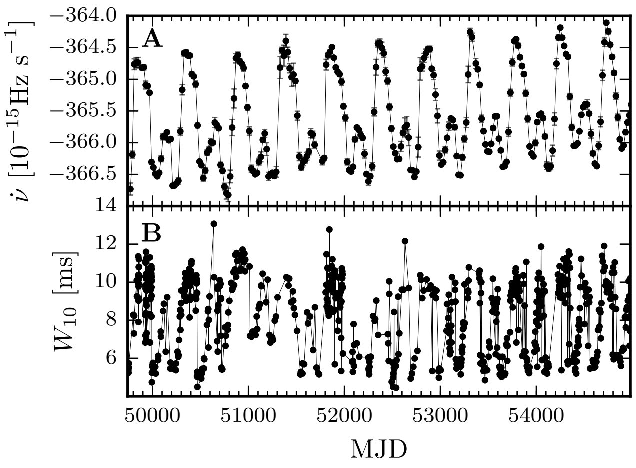

Figure 1. Observed data for PSR B1828-11 spanning from MJD 49710 to MJD 54980. In panel **A** we reproduce the spin-down rate with error-bars and in panel **B** the beam-width W_{10} (for which no error bars were available). All data courtesy of Lyne et al. (2010).

gular momentum vector; we will not consider such a model here, but note it is a valid alternative which deserves testing.

The distinctive spin-down of B1828-11 was analysed by Seymour & Lorimer (2013) for evidence of chaotic behaviour. They found evidence that B1828-11 was subject to three dynamic equations with the spin-down rate being one governing variable. This further motivates the precession model since it results from applying Euler's three rigid body equations to a non-spherical body (Landau & Lifshitz 1969).

The precession hypothesis was challenged by Lyne et al. (2010) when reanalysing the data. They noted that in order to measure the spin-down and beam-shape with any accuracy required time averaging over periods $\sim$ 100 days, smoothing out any behaviour acting on this time-scale. Motivated by the intermittent pulsar B1931+24, they put forward the phenomenological hypothesis that instead the magnetosphere is undergoing periodic switching between (at least) two metastable states. Such switching would result in correlated changes in the beam-width and spin-down rate. They returned to the data and instead of studying a time-averaged beam-shape-parameter as done by Stairs et al. (2000), they instead considered the beam-width at 10% of the observed maximum W_{10}. This quantity is time-averaged, but only for each observation which lasts $\sim$ 1hr. This makes W_{10} insensitive to any changes which occur on time-scales longer than an hour. If the meta-stable states last longer than this, W_{10} will be able to resolve the switching. The relevant data was kindly supplied to us courtesy of Lyne et al. (2010), and is reproduced in Fig. 1. On the basis of these observations, Lyne et al. (2010) concluded that the individual measurements of W_{10} for B1828-11 did in fact appear to switch between distinct high and low values, as opposed to a smooth modulation between the values, with this switching coinciding with the periodic changes in the spin-down. On this basis they concluded that B1828-11 was undergoing periodic switching between two magnetospheric states.

In our view, it is not immediately clear by eye whether the data presented in Fig. 1 is sufficient to rule out or even favour either of the precession or switching interpretations. For this reason, in this work we develop a framework in which to evaluate models built from these concepts and argue their merits quantitatively using a Bayesian model comparison. We note that a distinction must be made between a conceptual idea, such as precession, and a particular predictive model built from it. As we will see, each concept can generate multiple models, and furthermore we could imagine using a combination of precession and switching, with the precession acting as the 'clock' that modulates the probability of the magnetosphere being in one state or the other, an idea developed by Jones (2012). The models considered here cover the precession and switching interpretations, but we do not claim the models to be the 'best' that these hypotheses could produce.

The rest of the work is organised as follows: in Sec. 2 we will describe the framework to fit and evaluate a given model, in Sec. 3 we will define and fit several predictive models from the conceptual ideas, and then in Sec. 4 we shall tabulate the results of the model comparison. Finally, the results are discussed in Sec. 5.

2 BAYESIAN METHODOLOGY

We now introduce a general methodology to compare and evaluate models for this form of data. The technique is well practised in this and other fields and so in this section we intend only to give a brief overview; for a more complete introduction to this subject see Jaynes (2003); Gelman et al. (2013); Sivia & Skilling (1996).

2.1 The odds-ratio and posterior probabilities

There are two issues that we wish to address. Firstly, given two models, how can one say which is preferred, and by what margin? Secondly, assuming a given model, what can be said of the probability distribution of the parameters that appear in that model?

We can address the first issue by making use of Bayes theorem for the probability of model $\mathcal{M}_i$ given some data:

$$P(\mathcal{M}_i|\text{data}) = P(\text{data}|\mathcal{M}_i)\frac{P(\mathcal{M}_i)}{P(\text{data})}. \tag{1}$$

The quantity $P(\text{data}|\mathcal{M}_i)$ in known as the *marginal likelihood* of model $\mathcal{M}_i$ given the data.

In general we cannot compute the probability given in equation (1) because we do not have an exhaustive set of models to calculate $P(\text{data})$. However, we can compare two models, say A and B, by calculation of their *odds-ratio*:

$$\mathcal{O} = \frac{P(\mathcal{M}_A|\text{data})}{P\mathcal{M}_B|(\text{data})} = \frac{P(\text{data}|\mathcal{M}_A)}{P(\text{data}|\mathcal{M}_B)}\frac{P(\mathcal{M}_A)}{P(\mathcal{M}_B)}. \tag{2}$$

In the rightmost expression, the first factor is the ratio of the marginal likelihoods (also known as the *Bayes factor*) which we will discuss shortly, while the final factor reflects our prior belief in the two models. If no strong preference exists for one over the other, we may take a non-informative approach and set this equal to unity. We will follow this approach in what follows below.

We need to find a way of computing the marginal likelihoods, $P(\text{data}|\mathcal{M}_i)$. To this end, consider a single model $\mathcal{M}_i$ with model parameters $\boldsymbol{\theta}$, and define $P(\text{data}|\boldsymbol{\theta}, \mathcal{M}_i)$ as the *likelihood function* and $P(\boldsymbol{\theta}|\mathcal{M}_i)$ as the *prior distribution* for the model parameters. We can then perform the necessary calculations by making use of

$$P(\text{data}|\mathcal{M}_i) = \int P(\text{data}|\boldsymbol{\theta}, \mathcal{M}_i)P(\boldsymbol{\theta}|\mathcal{M}_i)d\boldsymbol{\theta}. \quad (3)$$

The likelihood function can also be used to explore the second issue of interest, by calculating the *joint-probability distribution* for the model parameters, also known as the *posterior probability distribution*:

$$P(\boldsymbol{\theta}|\text{data}, \mathcal{M}_i) = \frac{P(\text{data}|\boldsymbol{\theta}, \mathcal{M}_i)P(\boldsymbol{\theta}|\mathcal{M}_i)}{P(\text{data}|\mathcal{M}_i)}. \quad (4)$$

Note that the marginal likelihood $P(\text{data}|\mathcal{M}_i)$ described above plays the role of a normalising factor in this equation.

In general the integrand of Eqn. (3) makes analytic, or even simple numeric integration difficult or impossible. This will be the case for the probability model that we will use and so instead must turn to sophisticated numerical methods. For this study we use Markov-Chain Monte-Carlo (MCMC) techniques which simulate the joint-posterior distribution for the model parameters up to the normalising constant

$$P(\boldsymbol{\theta}|\text{data}, \mathcal{M}_i) \propto P(\text{data}|\boldsymbol{\theta}, \mathcal{M}_i)P(\boldsymbol{\theta}|\mathcal{M}_i). \quad (5)$$

In particular we will use the Foreman-Mackey et al. (2013) implementation of the affine-invariant MCMC sampler (Goodman & Weare 2012) to approximate the posterior density of the model parameters. Further details of our MCMC calculations can be found in Appendix A.

Once we are satisfied that we have a good approximation for the joint-posterior density of the model parameters we discuss how to recover the normalising constant to calculate the odds-ratio in Sec. 4.

2.2 Signals in noise

We now need to build a statistical model to relate physical models for the spin-down and beam-width to the data observed in Fig. 1. To do this we will turn to a method widely used to search for deterministic signals in noise.

We assume our observed data y^{obs} is a sum of a stationary zero-mean Gaussian noise process $n(t, \sigma)$ (here σ is the standard deviation of the noise process) and a signal model $f(t|\mathcal{M}_j, \boldsymbol{\theta})$ (where $\boldsymbol{\theta}$ is a vector of the model parameters) such that

$$y^{\text{obs}}(t_i|\mathcal{M}_j, \boldsymbol{\theta}, \sigma) = f(t_i|\mathcal{M}_j, \boldsymbol{\theta}) + n(t_i, \sigma). \quad (6)$$

Given a particular signal model, subtracting the model from the data should, if the model and model parameters are correct, leave behind a Gaussian distributed residual - the noise. That is

$$y^{\text{obs}}(t_i|\mathcal{M}_j, \boldsymbol{\theta}, \sigma) - f(t_i|\mathcal{M}_j, \boldsymbol{\theta}) \sim N(0, \sigma). \quad (7)$$

The data, for either the spin-down or beam-width, consists of N observations (y_i^{obs}, t_i). For a single one of these observations, the probability distribution given the model and model parameters is

$$P(y_i^{\text{obs}}|\mathcal{M}_j, \boldsymbol{\theta}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-\left(f(t_i|\mathcal{M}_j, \boldsymbol{\theta}) - y_i\right)^2}{2\sigma^2}\right\}. \quad (8)$$

The likelihood is the product of the N probabilities

$$P(\mathbf{y}^{\text{obs}}|\mathcal{M}_j, \boldsymbol{\theta}, \sigma) = \prod_{i=1}^{N} P(y_i^{\text{obs}}|\mathcal{M}_j, \boldsymbol{\theta}, \sigma), \quad (9)$$

where $\mathbf{y}^{\text{obs}}$ denotes the vector of all the observed data.

In Sec. 3 we will define the physical models, $f(t|\mathcal{M}_j, \boldsymbol{\theta})$, for the precession and switching interpretations; for now we recognise that once defined, we may calculate the likelihood of the data under the model using Eqn. (9).

2.3 Choosing prior distributions

In the previous section we have developed the likelihood function $P(\text{data}|\boldsymbol{\theta}, \mathcal{M}_j)$ for any arbitrary model producing a deterministic signal $f(t_i|\mathcal{M}_j, \boldsymbol{\theta})$ in noise. To compare between particular models, using Eqn. (2), we must compute the marginal likelihood as defined in Eqn. (3) which requires a prior distribution $P(\boldsymbol{\theta}|\mathcal{M}_j)$.

The choice of prior distribution is important in a model comparison since it can potentially have a large impact on the resulting odds-ratio. In general we want to use astrophysically informed priors wherever possible, or suitable uninformative (but proper) priors otherwise. However, the switching model presents a particular challenge in this respect, as its switching parameters (cf. Sec. 3.2) are ad-hoc and purely phenomenological, and were initially informed by the same data we are trying to test the models on. It is therefore important to avoid potential circularity in properly assessing the prior volume of its parameter space, which affects the relevant "Occam factor" for this model (e.g. see MacKay (2003)).

To resolve this, we will make use of the availability of two different and independent data sets: the spin-down and the beam-width data. First, we will perform parameter estimation using the spin-down data with uniform priors based on crude estimates from the data. For the model parameters common to both the spin-down and beam-width models, we will use the posterior distributions from the spin-down data as prior distributions for the beam-width model. For the remaining beam-width parameters which are not common to both the spin-down and beam-width models we will use astrophysically-motivated priors. In this way we can do model comparisons based on the beam-width data using proper, physically motivated priors. In addition, this enforces consistency between the beam-width and spin-down solutions: for example constraining the two to be in phase.

An obvious alternative is to do the reverse and use the beam-width data to determine priors for the spin-down data. However, for both models, we found difficulties in obtaining good quality posteriors when conditioning on the beam-width data with uniform priors based on crude estimates. The posteriors, we found to be non-Gaussian and multimodal-modal. To deal with this we would need to use a more sophisticated methodology than that discussed in Sec. A. By contrast this is not the case when conditioning on the spin-down data first (results presented in Sec. 3). This is expected since, even by eye, we see that the spin-down data contains an easily visible 'signal', while the beam-width data is relatively 'noisy'. For this work we are primarily interested in laying out the framework to perform model comparisons and either method should suffice and give the same solution. For now then, we will use the more straight-forward method of using the spin-down data to set priors for the beam-width.

3 DEFINING AND FITTING THE MODELS

In this section we will take each conceptual idea (precession or switching) and define a predictive signal model $f(t|M,\boldsymbol{\theta})$. Each concept may motivate multiple signal models: already we have seen the extension to the original Lyne et al. (2010) switching model by Perera et al. (2015). In this work we do not aim to exhaust all known models and are well aware that more models exist that have not yet been considered.

For each concept, we will first discuss the theoretical model, then discuss the choice of priors and finally the resulting posterior and posterior-predictive checks. For both these concepts we build models for both the spin-down and beam-width using the former to inform the priors for the latter as described in Sec. 2.3. Model comparisons will be made on the beam-width data only. In addition to these two concepts, we will also consider a noise-only model for the beam-width data.

It is worth stating that by using the signals-in-noise statistical model, we do not make any assumptions on the cause of the noise other than requiring it to be stationary and Gaussian (cf. Jaynes (2003)). Given the uncertain physics of neutron stars and the measurement of pulses, it seems likely the noise will contain contributions both from the neutron star itself, and from the measurement process, with the former dominating. We will add a subscript to the noise component $\sigma_{[\dot{\nu},W_{10}]}$ to distinguish between the two data sources.

3.1 Noise-only model

3.1.1 Defining the Noise-only beam-width model

Before evaluating the precession and switching hypothesis, let us first consider a noise-only model. This will introduce some generic concepts and provide a benchmark against which to test other models. The noise-only model asserts that the beam-width data (as seen in panel B of Fig.1) does not contain any periodic modulation, but is the result of noise about a fixed beam-width: the signal model $f(t) = W_{10}$ is a constant.

We will not consider the spin-down data under such a hypothesis since it is the beam-width data alone that we will use to make model comparisons and it is clear by eye that such a model is incorrect.

3.1.2 Fitting the model to the beam-width data

For the noise-only model we have two parameters which require a prior: the constant beam-width W_{10} and the noise $\sigma_{W_{10}}$. For the beam-width we will set a prior using astrophysical data from the ATNF database (available at www.atnf.csiro.au/people/pulsar/psrcat, for a description see Manchester et al. (2005)). Considering the entire normal radio-pulse population, we find that the W_{10} beam-widths are, to a good approximation, ln-normally distributed. Using a maximum-likelihood estimate we find the ln-distribution to have mean 3.3 and standard deviation 0.8. For the noise parameter $\sigma_{W_{10}}$ we will use a prior Unif$(0, 5\times10^{-3})$ s based on a crude estimate from the data. We must be careful here as by doing this, we are in a sense using the data twice, but this will not introduce bias into the model comparison provided the same prior is applied for all beam-width models.

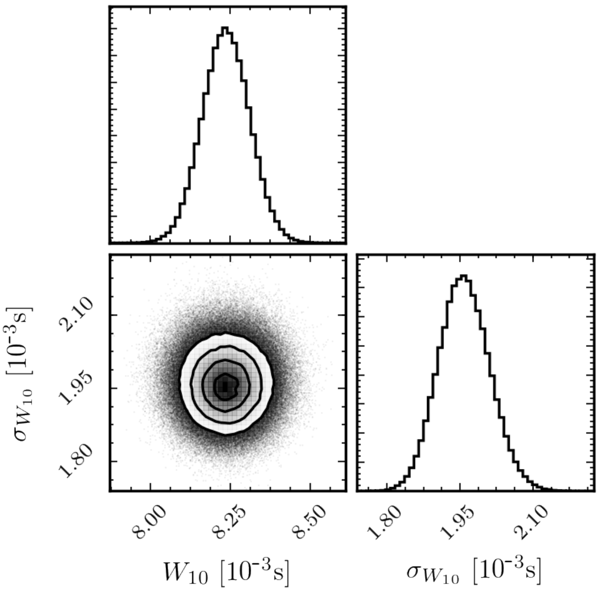

Figure 2. The estimated marginal posterior probability distributions for the Noise-only model parameters of the beam-width data.

The MCMC simulations converge quickly to a normal distribution as shown in Fig. 2. Of note is the mode of $\sigma_{W_{10}} \sim 2$ ms, this is the Gaussian noise required to explain the variations in W_{10} about a fixed mean. For other models, we hope to explain some of the variations with periodic modulation and the rest with Gaussian noise. So for these models we should expect $\sigma_{W_{10}} < 2$ ms.

In Fig. 3 we plot the *maximum posterior estimate* (MPE) of the signal alongside the data, i.e. the model prediction when the parameters are set equal to the peak values of the posterior probability distributions. This figure demonstrates that, for the noise-only model, the observed W_{10} has a mean value of approximately 8 ms, then all the variations about this mean are due to the noise. In the following section we will develop models where at least some the variation is explained by periodic modulations.

3.2 Switching model

The switching idea is phenomenological and we will build the model based on the modification of Lyne et al. (2010)

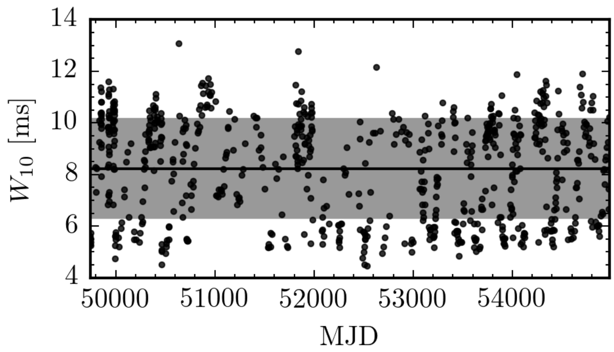

Figure 3. Posterior predictive check of the fit of the Noise-only model posterior distribution to the data: the solid black line is the maximum posterior estimate (MPE), i.e. the model prediction when the parameters are set equal to the values corresponding to the peaks of the posterior probability distributions. The shaded region indicates the MPE of the 1-$\sigma_{W_{10}}$ noise about the beam-width model, black dots are the original data.

by Perera et al. (2015): that is we assume the magnetosphere switches between two meta-stable states *twice* during a single period (the motivation for this is discussed in Sec. 3.2.1). For this work we will assume the switching to be deterministic, although improvements could be made by allowing the switching time to dither, or probabilistic variations in the switching states themselves; see Lyne et al. (2010) for some exploration of such ideas. This fully deterministic model captures the primary features without explaining the underlying physics, for example the cause of the switching. Both Jones (2012) and Cordes (2013) have worked to improve the physical motivations for the switching and provide a consistent picture. Nevertheless, in this work we choose to use the simple phenomenological model as a basis, which can be improved upon in future work.

3.2.1 Defining the spin-down rate model

The model proposed by Lyne et al. (2010) poses two states for the magnetosphere which we will label as S_1 and S_2. Then associated to each of these states is a corresponding spin-down rate $\dot{\nu}_1$ and $\dot{\nu}_2$. The smoothly varying spin-down that we observe is a result of the time-averaging process required to measure the spin-down rate. Lyne et al. (2010) suggested a square-wave-like switching with a duty rate measuring the fraction of time spent in one state compared to the other. They also proposed a dither in the switching period which will obscure the periodicity, and may give rise to low-frequency structure; we will not consider the dither in this work, but will investigate it in future work. While studying PSR B0919+06, which also demonstrates a double-peaked spin-down rate like B1828-11, the authors of Perera et al. (2015) realised that a (deterministic) switching model which flips once per cycle is incapable of explaining the double-peak observed in the spin-down rate (in particular that one peak is systematically smaller than the other). In order to explain this double-peaked structure, they propose that the mode-changes responsible for switching in the

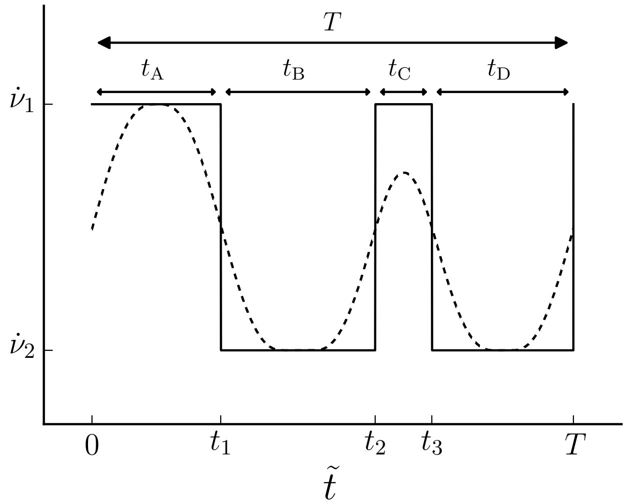

Figure 4. Schematic of the doubly-periodic spin-down rate model proposed by Perera et al. (2015). The solid line is the underlying spin-down evolution while the dashed line indicates the measured time-averaged quantity. In this instance, the time-average window is longer than $t_{\rm C}$, but shorter than the other three durations.

spin-down rate must be doubly-periodic: that is the spin-down rate changes state *twice* during a single cycle. Other modifications, such as introducing a third magnetospheric state, do exist, but in this work we will apply the Perera et al. (2015) switching model to B1828-11.

We will now discuss the particular formulation of this model used in this analysis, firstly defining the *underlying* spin-down model and then the time-averaging process. To aid in this discussion we plot both the underlying spin-down model and time-average in Fig. 4 and will gradually introduce each feature.

We begin by defining $\tilde{t} = (t - t_{\rm ref}) + \phi_0 T \mod (T)$ where $\phi_0 \in [0, 1]$ is an arbitrary phase offset and $t_{\rm ref}$ is a reference time. For all models in this study we will set $t_{\rm ref} =$ MJD 49621 to coincide with the epoch at which the ATNF database Manchester & Lyne (1977) records measurements for B1828-11. Then the function which generates the switching is

$$\dot{\nu}_{\rm P}(t) = \begin{cases} \dot{\nu}_1 & \text{if} \quad 0 < \tilde{t} < t_1 \text{ or } t_2 < \tilde{t} < t_3 \\ \dot{\nu}_2 & \text{if} \quad t_1 < \tilde{t} < t_2 \text{ or } t_3 < \tilde{t} < T \end{cases} . \tag{10}$$

There are multiple ways to parametrize the switching times in the model. For the data analysis, we have chosen to parametrize by the total cycle duration T and three of the segment durations $t_{\rm A}$, $t_{\rm B}$ and $t_{\rm C}$.

This model is subject to label-switching degeneracy in the choice of $\dot{\nu}_1$ and $\dot{\nu}_2$ and also between the various time-scales and initial phase ϕ_0. This degeneracy may cause difficulties in the MCMC search algorithm, and we therefore fix this gauge freedom by specifying that $|\dot{\nu}_1| < |\dot{\nu}_2|$, $t_{\rm A} \geq t_{\rm C}$,and we require that $t_{\rm A}$ refers to a segment where $\dot{\nu} = \dot{\nu}_1$.

Based on a cursory inspection of the observed B1828-11 spin-down rate (see Fig. 1 panel A), it is clear that the secular second-order spin-down rate is non-zero. To model this we will include a constant $\ddot{\nu}$ in the underlying spin-down model

$$\dot{\nu}(t) = \dot{\nu}_{\rm P}(t) + \ddot{\nu}(t - t_{\rm ref}) \, . \tag{11}$$

This gives the intrinsic spin-down rate of the pulsar which would be observed if measurements could be taken without time-averaging

To simulate the observed spin-down rate, we could time-average Eqn. (11) directly. Instead we choose to mimic the data collection process responsible for the time-averaging process. Let us first discuss the data collection process as described by Lyne et al. (2010). Observers start with the time-of-arrival of pulsations, which is a measure of the pulsar rotational phase. Taking a 100 day window of data, starting at the earliest observation, a second-order Taylor expansion in the phase is fitted to the data yielding a measurement of $\dot{\nu}$. Then the processes is repeated, sliding the window in intervals of 25 days over the whole data set. The measured $\dot{\nu}$ values at the centre of each window gives a time-averaged spin-down rate.

To mimic this data collection process, we first integrate Eqn. (11) twice to generate the phase and then repeat the above process. When integrating, we can ignore the arbitrary phase and frequency offsets since we discard them when calculating the spin-down rate. The resulting spin-down rates constitutes our signal model which is the time-average of Eqn. (11). A schematic representation of the sort of spin-down that is then found is given by the dotted curve in Figure Fig. 4. Clearly, the time-averaged spin-down is much smoother than than the underlying spin-down.

It is worth taking a moment to realise that the relation of the time-average spin-down to the underlying model $\dot{\nu}$ depends on both the segment durations and the length of the time average ($t_{\rm ave} = 100$ days). For the i^{th} segment, if the duration $t_i > t_{\rm ave}$ then the time-averaged spin-down will "saturate" and have a flat spot as in segment A of the illustration in Fig. 4. On the other hand, if $t_i < t_{\rm ave}$ then the maximum spin-down rate in this segment will be a weighted sum of the two underlying spin-down rates as in segment B of the illustration. The weighting is determined by the amount of time the underlying spin-down rate spends in each state during the time-average window.

3.2.2 *Parameter estimation for the spin-down*

Using crude estimates of the data on panel A of Fig. 1 we define uniform priors for the spin-down model parameters in Table 1. As previously mentioned, this means we are using the data twice: once in setting up the priors and once for the fitting. Therefore it would be inappropriate to use the results in a model comparison and this is not our intention: we want to use the posterior distribution as a prior for the beam-width parameter estimation.

For the spin-down data under the Switching model, the MCMC simulations converge quickly to a unimodal and normal-like distribution. The distribution is plotted in Fig. 5A, and we summarise the results by their mean and standard-deviation in Table 2.

To check that our fit is sensible, we plot the observed spin-down data in Fig. 5B alongside the maximum posterior estimate for the signal. The relative size of the noise component informs us how well the model fits the data: if $\sigma_{\dot{\nu}}$ is of a similar size to the variations in spin-down rate, then the model does poorly and we require a large noise component. In this case the noise-component is smaller than the variations in spin-down rate and the signal model explains most of the variations in the data.

Table 1. Prior distributions for the spin-down Switching model.

Parameter	Distribution	Units
$\dot{\nu}_1$	Unif(-3.66 × 10^{-13}, -3.64 × 10^{-13})	s^{-2}
$\dot{\nu}_2$	Unif(-3.67 × 10^{-13}, -3.66 × 10^{-13})	s^{-2}
$\ddot{\nu}$	Unif(0, 10 × 10^{-23})	s^{-3}
T	Unif(450, 550)	days
$t_{\rm A}$	Unif(0, 250)	days
$t_{\rm B}$	Unif(0, 250)	days
$t_{\rm C}$	Unif(0, 250)	days
ϕ_0	Unif(0, 1)	
$\sigma_{\dot{\nu}}$	Unif(0, 1 × 10^{-15})	s^{-2}

Table 2. Posterior estimates for the spin-down Switching model.

Parameter	Mean ± s.d.	Units
$\dot{\nu}_1$	$-3.65 \times 10^{-13} \pm 8 \times 10^{-17}$	s^{-2}
$\dot{\nu}_2$	$-3.66 \times 10^{-13} \pm 6 \times 10^{-17}$	s^{-2}
$\ddot{\nu}$	$8.89 \times 10^{-25} \pm 2 \times 10^{-25}$	s^{-3}
T	486.0±0.88	days
$t_{\rm A}$	158.0±7.85	days
$t_{\rm B}$	159.0±13.61	days
$t_{\rm C}$	15.55±4.40	days
ϕ_0	0.53±0.01	
$\sigma_{\dot{\nu}}$	$4.1 \times 10^{-16} \pm 2 \times 10^{-17}$	s^{-2}

Comparing the maximum posterior values of the four segment times to the time-average baseline (fixed at 100 days) can give an insight into how the model has best fit the data. If we take the posterior of $t_{\rm C}$, we find it has a mean value of ~ 15 days which is significantly shorter than the time-averaging baseline. For the other three segments, their durations are longer than this baseline. The reason that the fit in Fig. 5B has one maxima smaller than the other, is because the segment duration for that segment, $t_{\rm C}$, is shorter than the time-average baseline. This is expected and was precisely the motivation for using the model proposed by Perera et al. (2015); a switching model split into only two segments could not produce this feature.

On the other hand, we see in Fig. 5B that the model systematically fails to fit the second (slightly lower) *minima* of the doubly-peaked spin-down. This lack of fit suggests that the model may be lacking degrees of freedom and highlights an area of potential improvement.

3.2.3 *Defining the beam-width model*

For the beam-width, we assume that changes in the spin-down rate directly correlate to changes in this beam-width through changes in the beam geometry. Since we require the switching to be doubly-periodic for the spin-down to make sense, so to must we require the beam-width to be doubly periodic. That is we define the beam-width model to be

$$W_{\rm P}(t) \begin{cases} W_1 & \text{if} \quad 0 < \tilde{t} < t_1 \text{ or } t_2 < \tilde{t} < t_3 \\ W_2 & \text{if} \quad t_1 < \tilde{t} < t_2 \text{ or } t_3 < \tilde{t} < T \end{cases} . \tag{12}$$

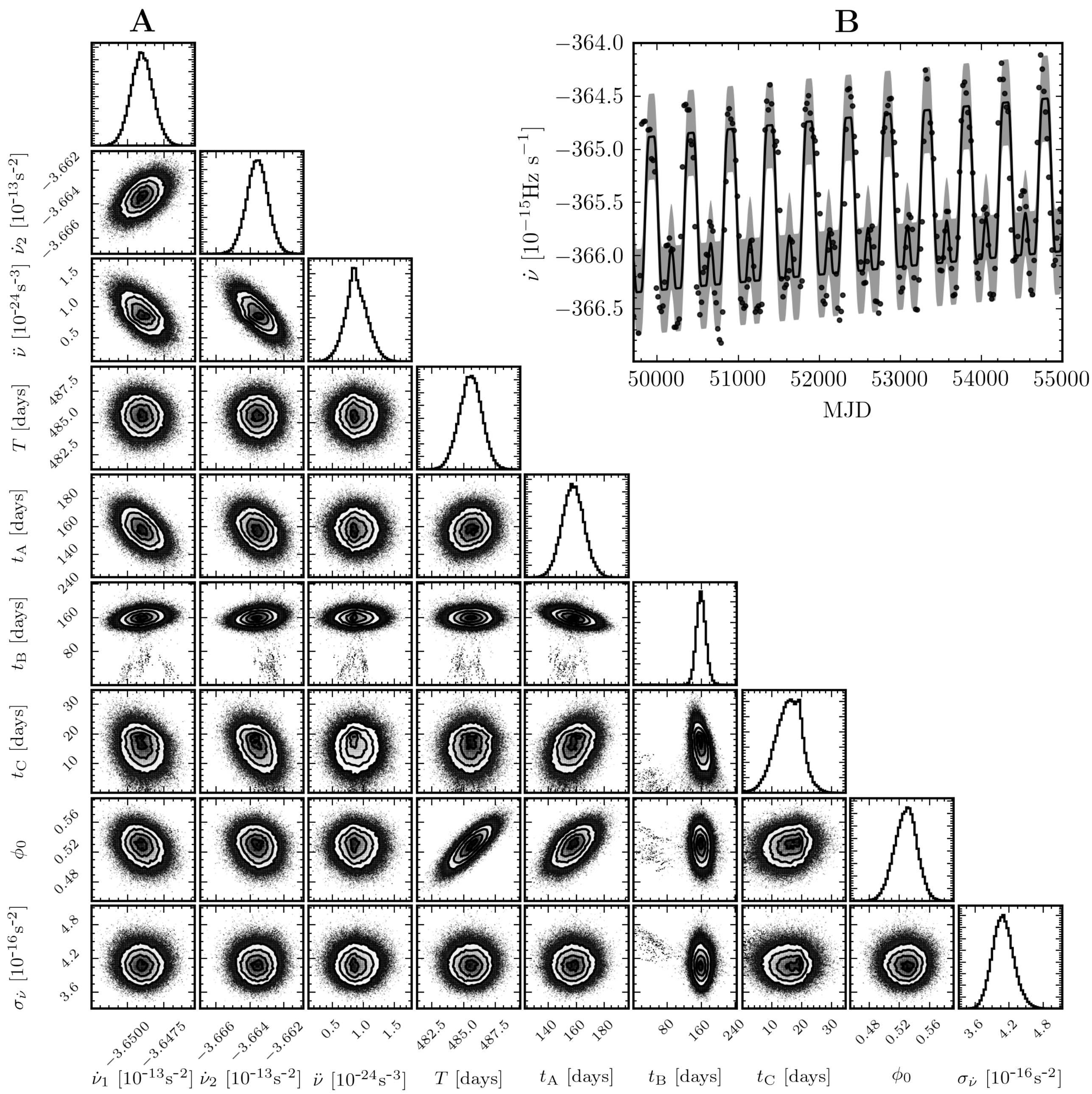

Figure 5. **A**: The estimated marginal posterior probability distribution for the Switching spin-down model parameters. **B**: Checking the fit of the model using the maximum posterior values to the data, see Fig. 3 for a complete description.

Lyne et al. (2010) noted that the larger beam-widths tended to correlate with the lower (absolute) spin-down rate ($\dot{\nu}_1$ in our model). We could fix this by requiring that $W_1 > W_2$ (recalling that we set $|\dot{\nu}_1| < |\dot{\nu}_2|$), but instead we will not implement such a constraint and allow the data to decide. As with the spin-down, to break the degeneracy in the times we will require again that $t_{\rm A} \geq t_{\rm C}$. We will not assume any secular changes in the beam-width for simplicity.

3.2.4 Parameter estimation for the beam-width

Having obtained a sensible fit to the spin-down data, we use the resulting posteriors (as summarised in Table 2) to inform our priors for the beam-width data. We can do this only for those parameters common to both the beam-width and spin-down predictions of the Switching model: namely the four time-scales, and the phase-offset. We would like to relate the spin-down rates $\dot{\nu}_1$ and $\dot{\nu}_2$ to the two beam-widths W_1 and W_2. However, the underlying physics is not understood, and so instead we will take a naive approach

Table 3. Prior distributions for the beam-width Switching model. Parameters for which the prior is taken from spin-down posteriors are labelled by *.

Parameter	Distribution	Units
W_1	Log$\mathcal{N}$(-3.30, 0.80)	s
W_2	Log$\mathcal{N}$(-3.30, 0.80)	s
T^*	$\mathcal{N}$(485.5, 0.88)	days
$t_A{}^*$	$\mathcal{N}$(158.0, 7.85)	days
$t_B{}^*$	$\mathcal{N}$(159.0, 13.61)	days
$t_C{}^*$	$\mathcal{N}$(15.55, 4.40)	days
$\phi_0{}^*$	$\mathcal{N}$(0.53, 0.01)	
$\sigma_{W_{10}}$	Unif(0, 5.0×10^{-3})	s

Table 4. Posterior estimates for the beam-width Switching model.

Parameter	Mean ± s.d.	Units
W_1	$9.51 \times 10^{-2} \pm 9 \times 10^{-5}$	s
W_2	$7.23 \times 10^{-3} \pm 8 \times 10^{-5}$	s
T	485.0±0.73	days
t_A	151.0±4.45	days
t_B	155.0±3.22	days
t_C	16.91±4.50	days
ϕ_0	$0.54 \pm 9 \times 10^{-3}$	
$\sigma_{W_{10}}$	$1.6 \times 10^{-3} \pm 4 \times 10^{-5}$	s

and set a prior on the beam-widths from astrophysical data. We already defined this prior in Sec. 3.1 for the beam-width in the Noise-only model, we will apply the same prior to both W_1 and W_2. The final parameter which requires a prior distribution is the noise-component: as described in Sec. 3.1 we apply a prior to $\sigma_{W_{10}}$ using a crude estimate from the data, this is tabulated along with the other priors in Table 3.

We now proceed to simulate the posterior, the posterior estimate is given in Fig. 6A and demonstrates non-Gaussianity and multimodal features in the segment times and the phase. This indicates the existence of multiple solutions which could explain the data. We note that the noise component $\sigma_{W_{10}}$ has a mode at 1.6 ms which is less than the 2ms required in the noise-only model. This indicates that some of the variability is being explained by the signal model. In Table 4 we summarise the posterior. We find that the posterior modes satisfy $W_1 > W_2$ which means that larger beam-widths are associated with the smaller absolute spin-down rates as found by Lyne et al. (2010).

Again we check the predictive power of our estimated posterior by plotting the MPE alongside the data in Fig. 6B. The fit to the data is not as good as the spin-down fit: by eye it is clear that most data points lie away from the signal model requiring a greater (relative) level of noise.

3.3 Precession model

We will now define the precession model and its predictions for the expected signal in the spin-down and beam-width data.

Classical free precession refers to the rotation of a rigid non-spherical body when there is a misalignment between its spin and rotation axes. For this work we will consider a biaxial star, acted upon by the vacuum dipole torque (Davis & Goldstein 1970). Such a star was considered by Jones & Andersson (2001), Link & Epstein (2001) & Akgün et al. (2006) in the context of B1828-11. Arzamasskiy et al. (2015) improved these models to include a plasma-filled magnetosphere; for illustration we will retain the vacuum approximation, but note this is an area of potential improvement. In the following we will work with the angles defined in Jones & Andersson (2001): that is the star emits EM radiation along the magnetic dipole $\mathbf{m}$ which makes an angle χ with the symmetry axis of the moment of inertia, and θ is the so-called wobble angle made between the symmetry axis and the angular momentum vector. We will consider the small-wobble angle regime where $\theta \ll 1$ since this is thought to be the most physical solution for B1828-11. Finally we define P as the rotation period, and $\dot{P}$ its time derivative, where the small variations due to precession have been averaged over, and

$$\tau_{\rm Age} \equiv \frac{P}{\dot{P}} \tag{13}$$

as the characteristic spin-down age[1] where I_0 is the moment of inertia of the star.

3.3.1 *Defining the spin-down rate model*

Observers infer the spin-down rate by measuring the arrival times of pulsations. For a freely precessing star the spin-down rate is periodically modulated on a time-scale known as the *free precession period*, which we will denote as $\tau_{\rm P}$. This result is referred to as the *geometric modulation* (Jones & Andersson 2001) since it is a geometric effect. Under the action of a torque the geometric effect persists, but an additional electromagnetic effect enters owing to torque variations (Cordes 1993). The authors of Jones & Andersson (2001) and Link & Epstein (2001) studied both effects and agreed that the electromagnetic contributions dominate for B1828-11. We will therefore neglect the geometric effect and calculate the spin-down rate to be

$$\dot{\nu}(t) = \frac{1}{2\pi\tau_{\rm Age}P}\left(-1 + 3\frac{1}{\tau_{\rm Age}}(t - t_{\rm ref}) + \theta\left[2\cot\chi\sin\left(\psi(t)\right) - \frac{\theta}{2}\cos\left(2\psi(t)\right)\right]\right), \tag{14}$$

where

$$\psi(t) = 2\pi\frac{t - t_{\rm ref}}{\tau_{\rm P}} + \psi_0. \tag{15}$$

As in the Switching model, $t_{\rm ref}$ = MJD 49621. The first two terms are the secular spin-down rate and its first derivative. The term in the square brackets is the modulation and can be found from either the sum of Eqn. (58) and (73) in Jones & Andersson (2001), or Eqn. (20) in Link & Epstein (2001) aside from a factor of χ in the harmonic term which we believe to be a misprint.

For $\chi < \pi/2$ the spin-down rate modulations are sinusoidal. When $\chi \approx \pi/2$ (such that the star is nearly an orthogonal rotator), we will see a strong harmonic at twice the precession frequency. It is precisely this behaviour which is

[1] Note that this is twice the characteristic spin-down age defined by Link & Epstein (2001)

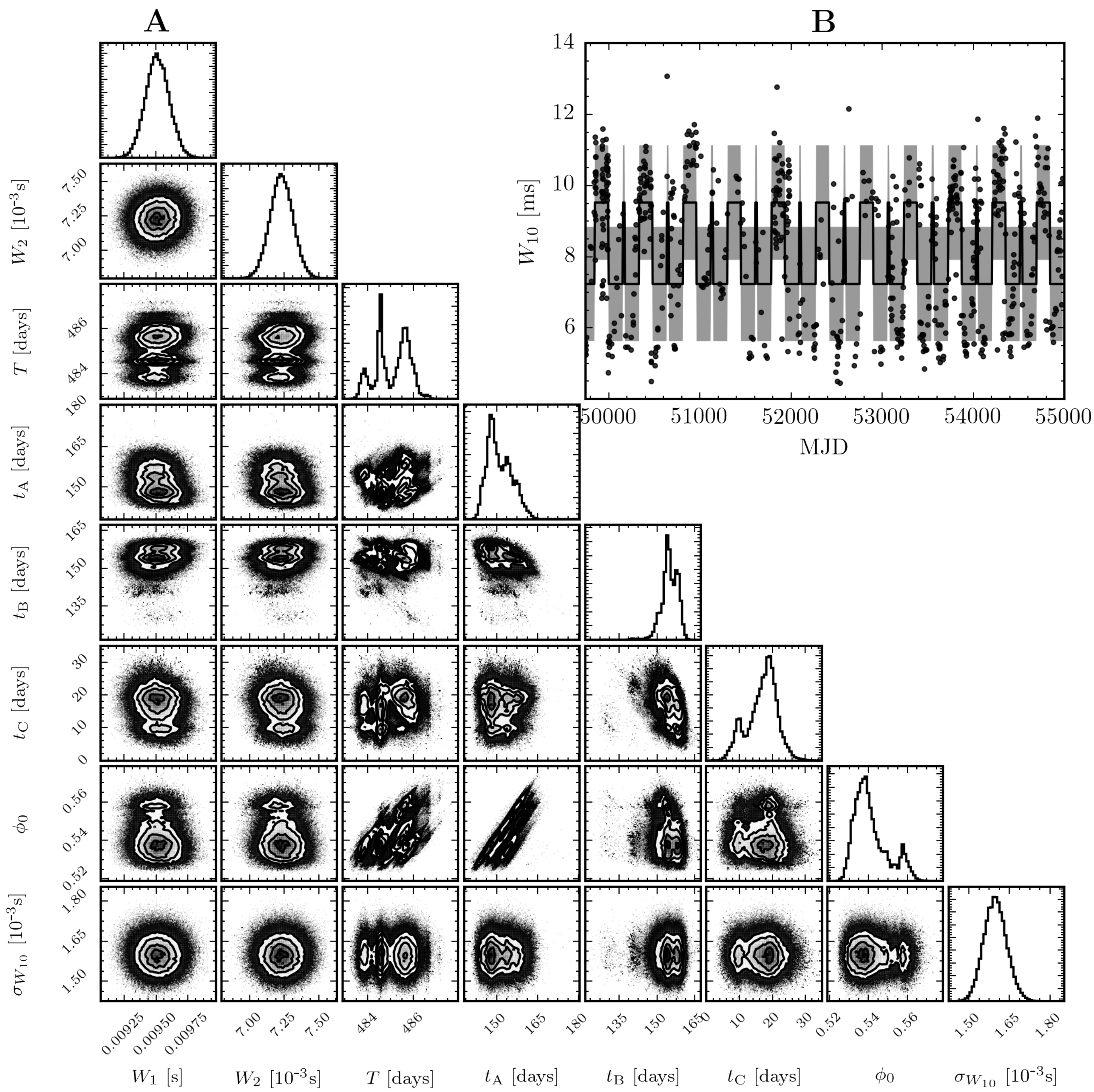

Figure 6. A: The estimated marginal posterior probability distribution for the Switching beam-width model parameters. **B**: Checking the fit of the model using the maximum posterior values to the data, see Fig. 3 for a complete description.

able to explain the doubly-peaked spin-down rate for B1828-11.

If fitted to the data using MCMC methods, the model proposed in Eqn. (14) suffers from strong dependency between P and $\tau_{\rm Age}$ which will lead to inefficient exploration of the parameter space. To prevent this, we will instead transform to the variables $\tau_{\rm Age}$ and $\zeta = \tau_{\rm Age}P$. We can then recover the distribution for P (which we require to set a prior for the beam-width inference) by looking at $\zeta/\tau_{\rm Age}$. It is worth stating that the period P is well known for B1828-11 and so we could alternatively use an astrophysical prior. The decision not to do this means that we will instead estimate P from the spin-down data. The results of this will be discussed in the next section.

3.3.2 Parameter estimation for the spin-down

In Table 5 we list the priors selected for the spin-down precession model. For $\tau_{\rm Age}, \zeta, \tau_{\rm P}$ and $\sigma_{\dot{\nu}}$ we use a prior based on a crude estimate from the data and for ψ_0 we give the full

Table 5. Prior distributions for the spin-down Precession model.

Parameter	Distribution	Units
$\tau_{\rm Age}$	Unif(0, 3.169×10^5)	yrs
ζ	Unif(0, 1×10^{13})	s^2
τ_P	Unif(450, 550)	days
θ	Unif(0, 0.1)	rad
χ	Unif($2\pi/5$, $\pi/2$)	rad
ψ_0	Unif(0, 2π)	rad
$\sigma_{\dot{\nu}}$	Unif(0, 1×10^{-15})	s^{-2}

Table 6. Posterior estimates for the spin-down Precession model.

Parameter	Mean ± s.d.	Units
$\tau_{\rm Age}$	$4.68 \times 10^4 \pm 2 \times 10^4$	yrs
ζ	$4.35 \times 10^{11} \pm 7 \times 10^7$	s^2
τ_P	486.0 ± 0.81	days
θ	0.049 ± 0.002	rad
χ	1.552 ± 0.001	rad
ψ_0	3.871 ± 0.069	rad
$\sigma_{\dot{\nu}}$	$4.04 \times 10^{-16} \pm 2 \times 10^{-17}$	s^{-2}

domain of possible values. Since our derivation of the signal models in Sec. 3.3 assumed the small wobble-angle regime $\theta \ll \chi$ and $\chi \sim \pi/2$, we similarly restrict their uniform priors to the relevant range.

Running the MCMC simulations we find fast convergence for all parameters although there is some noise in the initialisation run. We plot the resulting posterior in Fig. 7 and provide a summary in Table 6; for all parameters the distribution is Gaussian, except for $\tau_{\rm Age}$ which demonstrates a long tail. Nevertheless, $\tau_{\rm Age}$ is not used in the beam-width model and so the presence of the tail is not important for our model comparison.

In Fig. 7B we check the fit of the posterior for the spin-down data. The spin-down model fits to the data points well with only a small amount of noise required to explain the data. We note that the Precession model fails in much the same way as the Switching model in systematically overestimating the lower second minima. We discuss in Appendix C that this is a feature of the spin-down model, as given in Eqn. (14), but could suggest rethinking the role of the geometric modulations which are thought to be negligible.

As noted above, in order to avoid degeneracy during the MCMC simulations we used the model parameter $\zeta = \tau_{\rm Age} P$. For the beam-width data we require the distribution for P, the spin-period. We recover the distribution for P by looking at $\zeta/\tau_{\rm Age}$. In Fig. 8 we plot the distribution along with a normal distribution fitted using matching of moments.

We highlight here that the spin-period P is, of course, a well known quantity: the ATNF database (Manchester et al. 2005) provides a measurement of $0.405 \pm 1.2 \times 10^{-11}$ s. We would have been well justified in using such a measurement to inform our prior, or, since the error is so small, we could even justify using a fixed value for P. By instead including it as an unknown parameter, we have attempted to estimate P from the spin-down data alone. The resulting estimate can be compared with the known value in Fig. 8 and, given that we are not directly resolving individual pulses, provides a reasonable level of agreement at a fractional error of 0.21. We also ran the analysis using a fixed value of P, which resulted in no essential changes in the posterior estimates for the other parameters (or the odds-ratio to be presented later).

The posterior distributions conditioned on the spin-down data (as summarised in Table 6) can be compared with the values reported in Table 2 of Link & Epstein (2001). When comparing it should be noted that we are considering a longer stretch of data which includes most, but not all of the period studied by Link & Epstein (2001). For the two angles χ and θ, the fractional difference is 0.001 and 0.14 respectively while the precession periods differ by a fractional amount 0.05. These discrepancies reflect differences in the data span, but clearly the solution found here is similar to that found by Link & Epstein (2001).

3.3.3 Defining the beam-width model: Gaussian intensity

Modulation of the observed beam due to precession is a purely geometric effect. Fixing the beam-axis to coincide with the magnetic dipole $\mathbf{m}$ and following Jones & Andersson (2001), we define Θ and Φ as the polar and azimuthal angles of $\mathbf{m}$ with respect to a fixed Cartesian coordinate system with z along the angular momentum vector $\mathbf{J}$ and the observer in the x-z plane. The slow precessional motion of the spin-vector causes modulation in the angle Θ:

$$\Theta(t) = \cos^{-1}\left(\sin\theta \sin\chi \sin\psi(t) + \cos\theta\cos\chi\right), \tag{16}$$

which, in the $\theta \ll 1$ limit is approximately

$$\Theta(t) \approx \chi - \theta \sin\psi(t). \tag{17}$$

Taking the plane containing the angular momentum vector and the observer, in Fig. 9 we demonstrate the range of motion of $\mathbf{m}$ over a precessional cycle by a gray shaded region. The region has a mean polar value of χ and a range of 2θ.

For an observer fixed at an angle ι to $\mathbf{J}$ we define an impact parameter:

$$\Delta\Theta(t) = \Theta(t) - \iota, \tag{18}$$

which will vary in time with the precession period $\tau_{\rm P}$. This impact parameter determines how the observer's line-of-sight cuts the emission beam; if the emission varies due to changes in $\Delta\Theta(t)$, then the observer will measure a time-varying emission. To help visualise the setup, in Fig. 10 we plot the unit sphere with points corresponding to the beam-axis $\mathbf{m}$ and the observer. For each of these we have added lines of latitude and longitude. Then we see that that $\Delta\Theta$ is the difference between the lines of latitude and we can also define $\Delta\Phi(t) = \Phi(t) - \Phi_{\rm obs}$ as the difference in the lines of longitude.

The analysis by Stairs et al. (2000) characterised the beam by a shape parameter. Link & Epstein (2001) used an expansion in $\Delta\Theta$ to model the beam-width and hence the shape-parameter. This allowed them to use their fit to the timing-data to infer the beam-geometry which they found to be hour-glass shaped; see their Figure 5 for a schematic illustration. The authors of Lyne et al. (2010) did not use a shape-parameter as it requires time-averaging over a longer base-line, something they wish to avoid in order to be able to observe the switching. Instead, they considered the beam-width at 10% of the maximum, W_{10}, which is measured on a

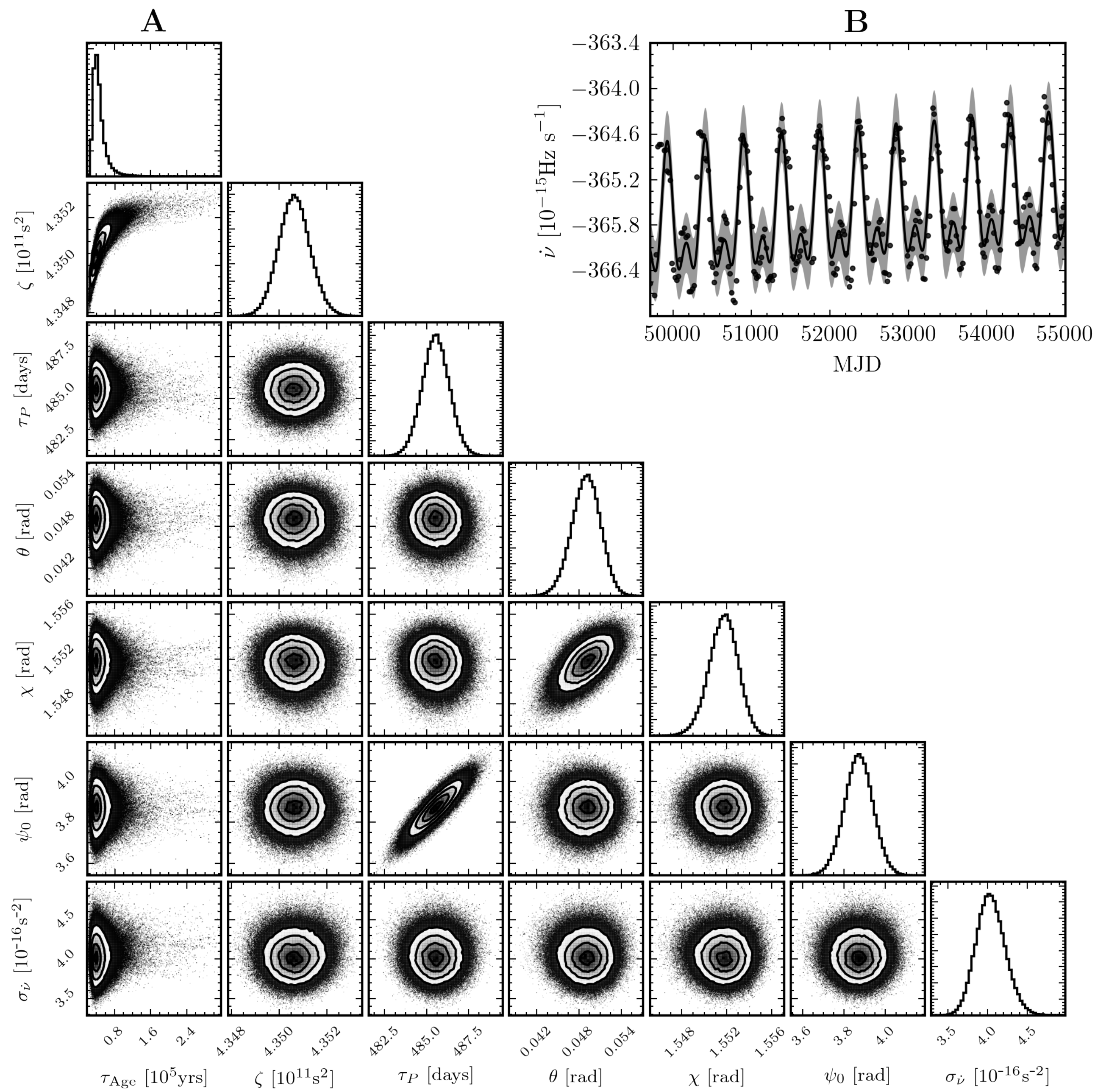

Figure 7. A: The estimated marginal posterior probability distribution for the Precession spin-down model parameters. **B**: Checking the fit of the model using the maximum posterior values to the data, see Fig. 3 for a complete description.

shorter time-baseline (~ 1 hr). If we want to use the beam-width to make a model comparison, we will require a model for W_{10} that is not informed by the data.

The integrated pulse profile of B1828-11 (Fig. 4 of Lyne et al. (2010)) shows a single peak, often described as core emission (Lyne & Manchester 1988). Since we do not have a detailed model of the emission mechanism, we will now consider the most rudimentary and natural beam geometry which fits this: a circularly symmetric (about the beam-axis) intensity which falls-of with a Gaussian function. Specifically, let us define Δd as the central angle between the observer's line-of-sight and the beam (this is the spherical distance between the two points marked in Fig. 10), then the intensity is

$$\mathcal{I}(t) = \mathcal{I}_0 \exp\left(-\frac{\Delta d(t)^2}{2\rho^2}\right). \tag{19}$$

Here $\mathcal{I}_0$ is the intensity when observed directly along the dipole and ρ measures the angular width of the beam. From the spherical law of cosines, for an observer located

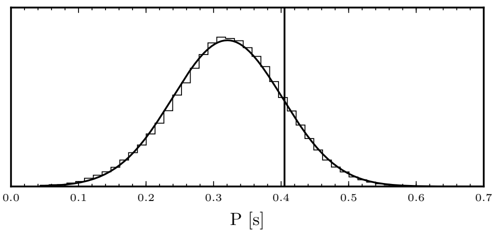

Figure 8. The posterior distribution for P in the spin-down precession model. The thin line is the binned histogram from the MCMC simulations, a thicker black line marks the normal distribution fitted with matched moments and a vertical line marks the known value of $P = 0.405$ s (Manchester et al. 2005).

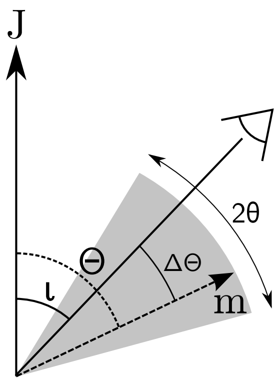

Figure 9. Illustration of the angles as the beam-axis **m** cuts the plane containing the observer and the angular momentum **J**.

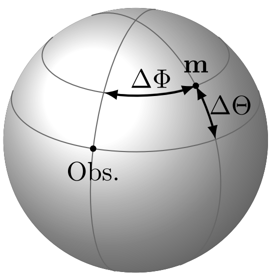

Figure 10. Illustration of the angles $\Delta\Theta$ and $\Delta\Phi$ relating the observer to the beam-axis **m**.

at $(\Phi_{\rm obs}, \iota)$, we have

$$\Delta d(t) = \cos^{-1}\left[\cos\Theta(t)\cos\iota + \sin\Theta(t)\sin\iota\cos|\Delta\Phi(t)|\right]. \tag{20}$$

The observer will see a maximum pulse intensity at $\Delta\Phi = 0$, given by

$$\mathcal{I}_{\rm max} = \mathcal{I}_0 \exp\left(-\frac{(\Theta(t) - \iota)^2}{2\rho^2}\right). \tag{21}$$

Now let us recognise that Θ varies on the slow precession time-scale, while Φ varies on the rapid spin time-scale: a single pulse consists of Φ varying between $\Phi_{\rm obs} - \pi$ and $\Phi_{\rm obs} + \pi$. So over a single pulse, we can treat Θ as a constant. The pulse width W_{10}, as measured by observers, is the duration for which the pulse intensity is greater than 10% of the peak observed intensity. For a single pulse, we can define this duration as the period for which the inequality

$$\mathcal{I} > \mathcal{I}_{\rm max}\frac{1}{10}, \tag{22}$$

is satisfied. Substituting equations (19) and (21) into (22) and rearranging we find that

$$\cos(|\Delta\Phi(t)|) > \frac{\cos\Psi(t) - \cos\Theta(t)\cos\iota}{\sin\Theta(t)\sin\iota}, \tag{23}$$

where

$$\Psi(t) = \sqrt{(\Theta(t) - \iota)^2 + 2\rho^2\ln(10)}. \tag{24}$$

Since we treat Θ as a constant over a single pulsation, we can also treat the whole right-hand-side of the inequality as a constant during each pulse.

Now consider a single rotation with the magnetic dipole starting and ending in the antipodal point to the observer's position such that $\Phi - \Phi_{\rm obs}$ increase between $-\pi$ and π during this rotation. Then the inequality of (23) measures the fraction of the pulse corresponding to the beam-width measurement. In terms of the rotation, we define $\delta\Phi$ as the angular width for which the inequality is satisfied and calculate it to be

$$\delta\Phi(t) = 2\cos^{-1}\left(\frac{\cos\Psi(t) - \cos\Theta\cos\iota}{\sin\Theta\sin\iota}\right). \tag{25}$$

Then the beam-width is

$$W_{10}(t) = P\frac{\delta\Phi(t)}{2\pi}, \tag{26}$$

from which we arrive at

$$W_{10}(t) = \frac{P}{\pi}\cos^{-1}\left(\frac{\cos\Psi(t) - \cos\Theta(t)\cos\iota}{\sin\Theta(t)\sin\iota}\right). \tag{27}$$

In order for the observer to measure the width at 10% of the maximum, the beam intensity must of course drop below this value before increasing again. In reality we typically observe pulse durations lasting for small fractions of the period, especially when they are close to orthogonal rotators (Lyne & Manchester 1988). We defined the intensity shape parameter ρ with units of radians, so a value of ~ 1 would imply a pulse duration lasting a significant fraction of the period. Since we believe it should be less than this, in this work then we will apply a prior of

$$P(\rho) \sim \text{Unif}(0, 0.1), \tag{28}$$

on the intensity-shape parameter.

3.3.4 *Parameter estimation for the Gaussian beam-width model*

We are in a position to fit the Gaussian beam model to the observed W_{10} values. In Table 7 we list the priors taken from the spin-down fit along with three additional priors. For ι we choose a uniform prior in $\cos\iota$ on $[-1, 1]$, this corresponds to allowing ι to range from $[0, \pi]$ (the observer could be in either hemisphere); for ρ we apply the prior from Eqn. (28); and for $\sigma_{W_{10}}$ we use a crude estimate based on the data (again we use the same prior for all models).

Fitting Eqn. (27) to the data we discover that the Gaussian beam model is a poor fit to the data. In Fig. 11B the MPE shows that while the model is able to fit the averaged

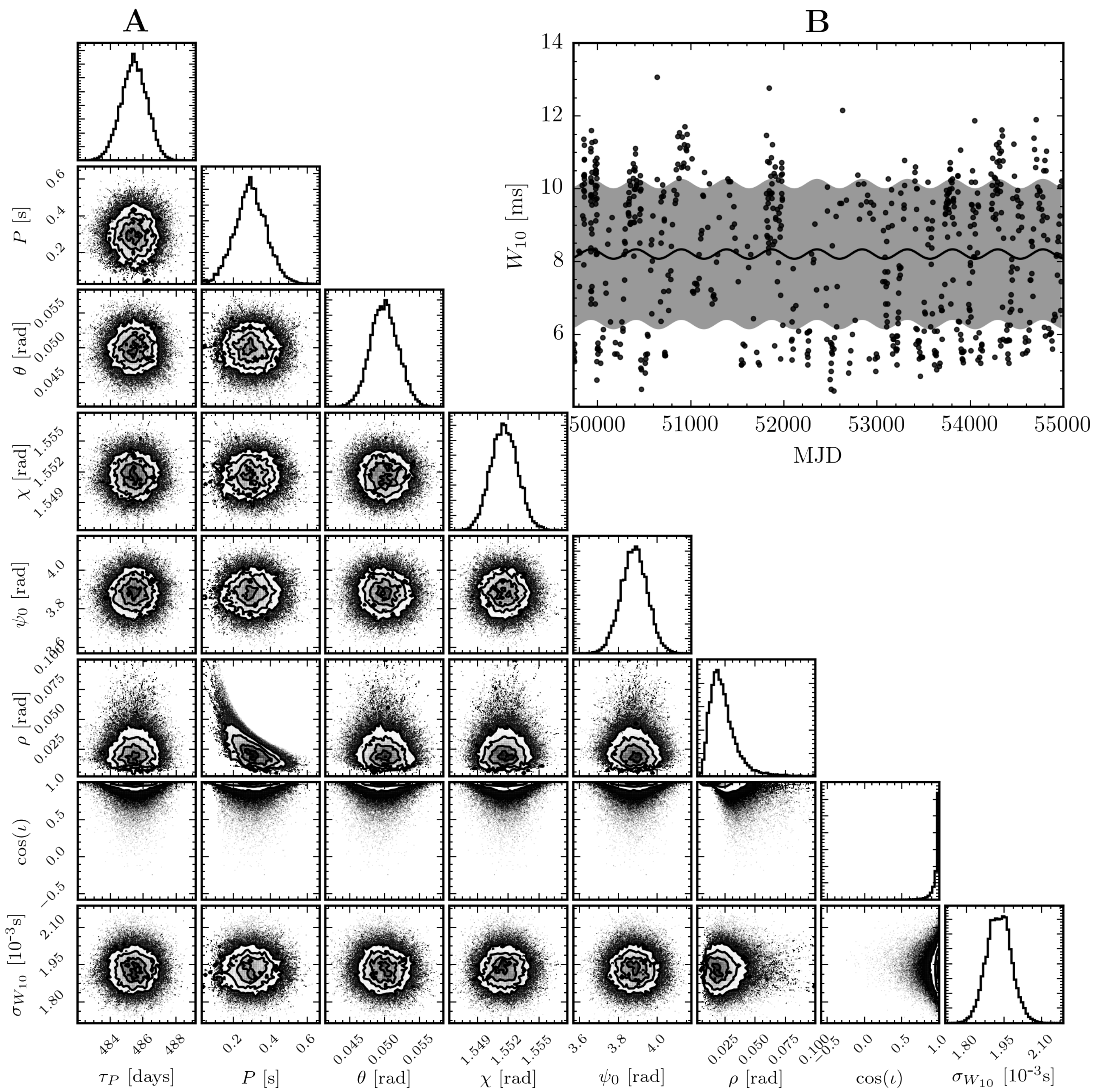

Figure 11. **A**: The estimated marginal posterior probability distribution for the Gaussian spin-down model parameters. **B**: Checking the fit of the model using the maximum posterior values to the data, see Fig. 3 for a complete description.

beam-width, it cannot simultaneously fit the amplitude of periodic modulations.

The posterior distribution (as seen in Fig. 11A) is Gaussian for all of the parameters except $\cos\iota$ for which it concentrates the probability at $\iota \approx 0$: the observer looks almost down the angular momentum vector. Since $\chi \approx \pi/2$ and $\theta \ll 1$, for each pulsation the beam sweeps out a cone with such a large opening angle it is close to a plane orthogonal to the rotation vector. Meanwhile the rotation vector is nearly parallel to the angular momentum, since $\theta \ll 1$. As a result, the beam remains approximately orthogonal to the observer for the entirety of each pulsation. We find this result difficult to believe on the grounds that the observer would *not* see tightly coliminated pulsed emission. For this reason, we conclude that the Gaussian beam-intensity fails to fit the data because the best-fit is unphysical. In retrospect, this result is not surprising since a Gaussian beam intensity is known to have a beam-width (as measured by W_{10}) which is independent of the impact angle as discussed by Akgün et al. (2006). This is a direct result of measuring the beam-width

Table 7. Prior distributions for the beam-width Gaussian Precession model. Parameters for which the prior is taken from spin-down posteriors are labelled by *.

Parameter	Distribution	Units
$\tau_P{}^*$	$\mathcal{N}(485.5,\, 0.81)$	days
P^*	$\mathcal{N}(0.32,\, 0.08)$	s
θ^*	$\mathcal{N}(0.049,\, 0.002)$	rad
χ^*	$\mathcal{N}(1.552,\, 0.001)$	rad
$\psi_0{}^*$	$\mathcal{N}(3.871,\, 0.069)$	rad
ρ	Unif(0, 0.1)	rad
$\cos(\iota)$	Unif(-1, 1)	
$\sigma_{W_{10}}$	Unif(0, 5.0×10^{-3})	s

with respect to the observed maximum and the self-similar nature of the Gaussian intensity under changes in the impact parameter.

3.3.5 *Refining the beam-width model: Modified-Gaussian intensity*

As we have demonstrated that the Gaussian beam is unable to explain both the observed variations and average beam-width we must now consider how it could be varied in a natural way which does explain the data. One suggestion from Akgün et al. (2006) is to impose a sharper cut-off, or introduce a conal component in addition to the Gaussian core emission. We will follow a slightly different path below, one which represents a less drastic modification of the beam profile.

The beam intensity described by Eqn. (19) is circularly symmetric about the beam axis as viewed on the surface of the sphere. In the context of the hollow-beam model (Radhakrishnan & Cooke 1969), Narayan & Vivekanand (1983) found that pulsar beams can be elongated with the ratio of major to minor axis being ~ 3 for typical pulsars. B1828-11 does not fit into the hollow-beam model (having only a single core component), but nevertheless if the conal emission can be non-circular a generalisation of our core intensity would be to allow for an elliptical beam.

To consider non-symmetric geometries, let us take the planar limit of Eqn. (20) by applying small angle approximations in $\Delta d, \Delta\Theta$ and $\Delta\Phi$:

$$\Delta d(t)^2 = \Delta\Theta(t)^2 + \sin\Theta(t)\sin\iota\Delta\Phi(t)^2. \tag{29}$$

This corresponds to setting the observer close to $\mathbf{m}$ in Fig. 10.

Obviously $\Delta\Phi$ ranges over $[0, 2\pi]$ in each rotation, but when $\Delta\Phi$ is not small, the intensity vanishes rapidly due to the Gaussian beam shape Eqn. (21). Therefore, Eqn. (29) is a good approximation for the separation when the beam is pointing near to the observer, while away from this it is a poor approximation, but the intensity is negligible and so the differences are inconsequential.

We can now allow for an elliptical beam geometry by postulating the beam intensity to be

$$\mathcal{I}(t) = \mathcal{I}_0 \exp\left(-\frac{-\Delta\Theta(t)^2}{2\rho_1^2} - \frac{(\sin\Theta(t)\sin\iota\Delta\Phi(t))^2}{2\rho_2^2}\right). \tag{30}$$

Then to calculate the beam-width, we first find the maximum:

$$\mathcal{I}_{\max} = \mathcal{I}_0 \exp\left(-\frac{-\Delta\Theta(t)^2}{2\rho_1^2}\right). \tag{31}$$

Solving for the beam-width we find

$$W_{10}(t) = \frac{P}{\pi}\frac{\sqrt{2\ln 10}\rho_2}{\sin\Theta(t)\sin\iota}, \tag{32}$$

which is independent of ρ_1, the latitudinal standard-deviation. The extra degree of freedom introduced in Eqn. (30) is irrelevant to the beam-width measure W_{10} because it is defined by the ratio of the intensity to that at the maximum $\mathcal{I}_{\max}$.

This loss of a degree of freedom means that Eqn. (32) is an equivalent to an expansion of Eqn. (27) in the planar limit (i.e. the non-circular nature introduced by Eqn. (30) does not manifest in the prediction for W_{10}) and so will suffer the same problems if fitted to the data. To further generalise our intensity model we will therefore modify the beam-geometry by allowing a varying degree of non-circularity. This is done by expanding the longitudinal standard deviation as

$$\rho_2(t) = \rho_2^0 + \rho_2''\Delta\Theta(t)^2. \tag{33}$$

Note that we have neglected to include a linear term here, forcing the geometry to be longitudinally symmetric about the beam-axis. Preliminary studies began by fitting a linear term only (this giving a modulation at the frequency $1/\tau_{\rm P}$), but it was found that including a second-order term (which provides modulation at both $1/\tau_{\rm P}$ and $2/\tau_{\rm P}$) gave a better fit. Including both terms, we found that the data was unable to provide inference on both ρ_2' and ρ_2'' due to degeneracy. In light of this, we drop the first term, but keep the second, which we feel is the simplest model which is able to fit the data.

Solving for the beam-width we obtain a signal model

$$W_{10}(t) = \frac{P}{\pi}\sqrt{\frac{2\ln 10}{\sin\Theta(t)\sin\iota}}\left(\rho_2^0 + \rho_2''\Delta\Theta(t)^2\right), \tag{34}$$

which we will refer to as the Precession beam-width model.

3.3.6 *Parameter estimation for the Modified-Gaussian Precession beam-width*

For equation (34), we define the prior distributions in table 8. As in the previous Gaussian model, we let ι range over $[0, \pi]$; for ρ_2^0, we apply the prior on intensity widths as given by Eqn. (28); and for ρ_2'' we will use a normal prior with zero mean favouring a Gaussian intensity. The *standard-deviation* of this prior can have a measurable impact on the inference: if it is too small then the degree of freedom introduced by Eqn. 33 is effectively removed. Instead we want to make it significantly larger than the (apriori unknown) posterior value of ρ_2'': this generates a so-called non-informative prior. To set the prior standard-deviation then, we need to provide a rough scale for what value ρ_2'' should have. To do this we will define our prior expectation as:

$$\rho_2(\Delta\Theta = \rho_2^0) \sim 2\rho_2^0, \tag{35}$$

which is to say we expect the ρ_2 to increase by no more than a factor of order unity over angular distances of the beam-width comparable to ρ_2^0 (the beam-width when the

Table 8. Prior distributions for the beam-width Modified-Gaussian Precession model. Parameters for which the prior is taken from spin-down posteriors are labelled by *.

Parameter	Distribution	Units
$\tau_P{}^*$	$\mathcal{N}(485.5, 0.81)$	days
P^*	$\mathcal{N}(0.32, 0.08)$	s
θ^*	$\mathcal{N}(0.049, 0.002)$	rad
χ^*	$\mathcal{N}(1.552, 0.001)$	rad
$\psi_0{}^*$	$\mathcal{N}(3.871, 0.069)$	rad
ρ_2	Unif(0, 0.1)	rad
ρ_2''	$\mathcal{N}(0, 10)$	rad^{-2}
$\cos(\iota)$	Unif(-1, 1)	
$\sigma_{W_{10}}$	Unif(0, 5.0×10^{-3})	s

Table 9. Posterior estimates for the beam-width Modified-Gaussian Precession model.

Parameter	Mean ± s.d.	Units
τ_P	485.0±0.47	days
P	0.29±0.08	s
θ	0.049±0.002	rad
χ	1.552±0.001	rad
ψ_0	3.970±0.040	rad
ρ_2	0.038±0.012	rad
ρ_2''	5.29±1.74	rad^{-2}
$\cos(\iota)$	$7.93\times10^{-3}\pm2\times10^{-3}$	
$\sigma_{W_{10}}$	$1.58\times10^{-3}\pm4\times10^{-5}$	s

observer cuts directly through the beam-axis). This amount to assuming that the beam does not depart very far from circularity. Given the prior on ρ_2^0 we set a standard-deviation of $\rho_2''=10$ which should cover the region of inference while remaining only weakly informative. We also tested different choices of ρ_2'' and found that the posteriors and odds-ratios where robust to the choice provided the standard-deviations was greater than unity.

The MCMC simulations converge quickly to a Gaussian distribution as shown in Fig. 12A. We observe mild dependency between the period P and the beam-geometry parameters ρ_2 and ρ_2'' though this did not slow down the convergence. The model parameters common to the spin-down model do not vary significantly from the spin-down posterior: this indicates the two models are consistent. We find that ι is close to $\pi/2$ as expected, ρ_2 is sufficiently small indicating a narrow pulse beam, but ρ_2'' has departed from its prior mean of zero. This confirms that our generalisation of the Gaussian intensity, Eqn. (33), is important in fitting the data. In Table 9 we summarise the posterior.

In Fig. 12B we perform the posterior predictive check plotting the MPE alongside the data. This demonstrates that the best-fit puts χ within θ of ι such that during the precessional cycle the beam-axis passes twice through observer's location. This corresponds to the gray region in Fig. 9 intersecting the observer's line-of-sight. When this happens, the modulation of the beam-width picks up a second harmonic at twice the precession frequency. The minima in Fig. 12B corresponds to the point in the precessional phase when the beam-axis point directly down the observer's line-of-sight.

3.3.7 Recreating the beam-geometry

Since we have defined a beam-intensity in Eqn. (30) we can recreate the beam-geometry and pulse-shape from our MPE values. The data we have does not provide information about the latitudinal beam-shape parameter ρ_1; therefore we consider that there are a family of beam-geometries parametrized by $\rho_1=\lambda\rho_2^0$ where λ is an arbitrary scale parameter and ρ_2^0 is the MPE value.

In figure 13 we pick four illustrative values for λ and plot the resulting beam-geometry as contour lines at fixed fractions of the maximum beam intensity (which occurs at the origin). This demonstrates that the beam-geometry has an hour-glass shape in agreement with Link & Epstein (2001), although this becomes weaker with smaller values for λ.

In Fig. 13, a pulse corresponds to a horizontal cut through the intensity at fixed $\Delta\Theta$. Our posteriors distribution, Fig. 12A, also provides information on how the observations cut through this beam-geometry. Under the precession hypothesis, the observer has a time-averaged $\Delta\Theta$ of $\chi-\iota$: this has been plotted as a horizontal dashed line in Fig. 13. Precession modulates $\Delta\Theta$ about this average value by $\pm\theta$; the observer's line-of-sight through the beam therefore varies by $2\theta\approx0.1$ rad over a precessional cycle. We have plotted a gray shaded region in Fig. 13 to show the extent, $\chi-\iota\pm\theta$, over which $\Delta\Theta$ varies during a precessional cycle.

We stress here that the contour lines cannot be used directly to measure the beam-width W_{10}. This is because W_{10} is defined as the width at 10% of the peak intensity for that observed pulse and not the maximum intensity of the beam. The peak intensity for an observed pulse (a horizontal slice) is the intensity at $\Delta\Phi=0$ and it is with respect to this, which W_{10} is measured.

By construction, the four beam geometries in Fig. 13 all produce the same W_{10} behaviour as observed in Fig. 12B. The reason for this is that we have lost information on the total intensity by using W_{10}; other beam-width measured could potentially yield more information and better constrain the beam geometry.

Fixing $\lambda=1$ we can also consider the variations in the pulse profile. In Fig. 14 we plot the normalised intensity for three values of $\Delta\Theta$ corresponding to the mean, and edges of the gray region in Fig. 13. This figure shows that the narrow beam-widths occur when $\Delta\Theta$ is small, which, since χ is close to $\pi/2$ coincide with the larger (absolute) spin-down rates. This agrees with the findings of Lyne et al. (2010) and this figure can be directly compared with panel C in Fig. 3 of that work.

4 ESTIMATING THE ODDS-RATIO

4.1 Thermodynamic integration

Having checked that our MCMC simulations are a reasonable approximation to the posterior distribution we now calculate the marginal likelihood for each model and then their odds-ratio. To calculate the marginal likelihood we will use thermodynamic integration. This requires running N parallel MCMC simulations and raising the likelihood to a power $1/T$ where T is the 'temperature' of the chain. This method was originally proposed by Swendsen & Wang (1986) to improve the efficiency of MCMC simulations for multimodal-

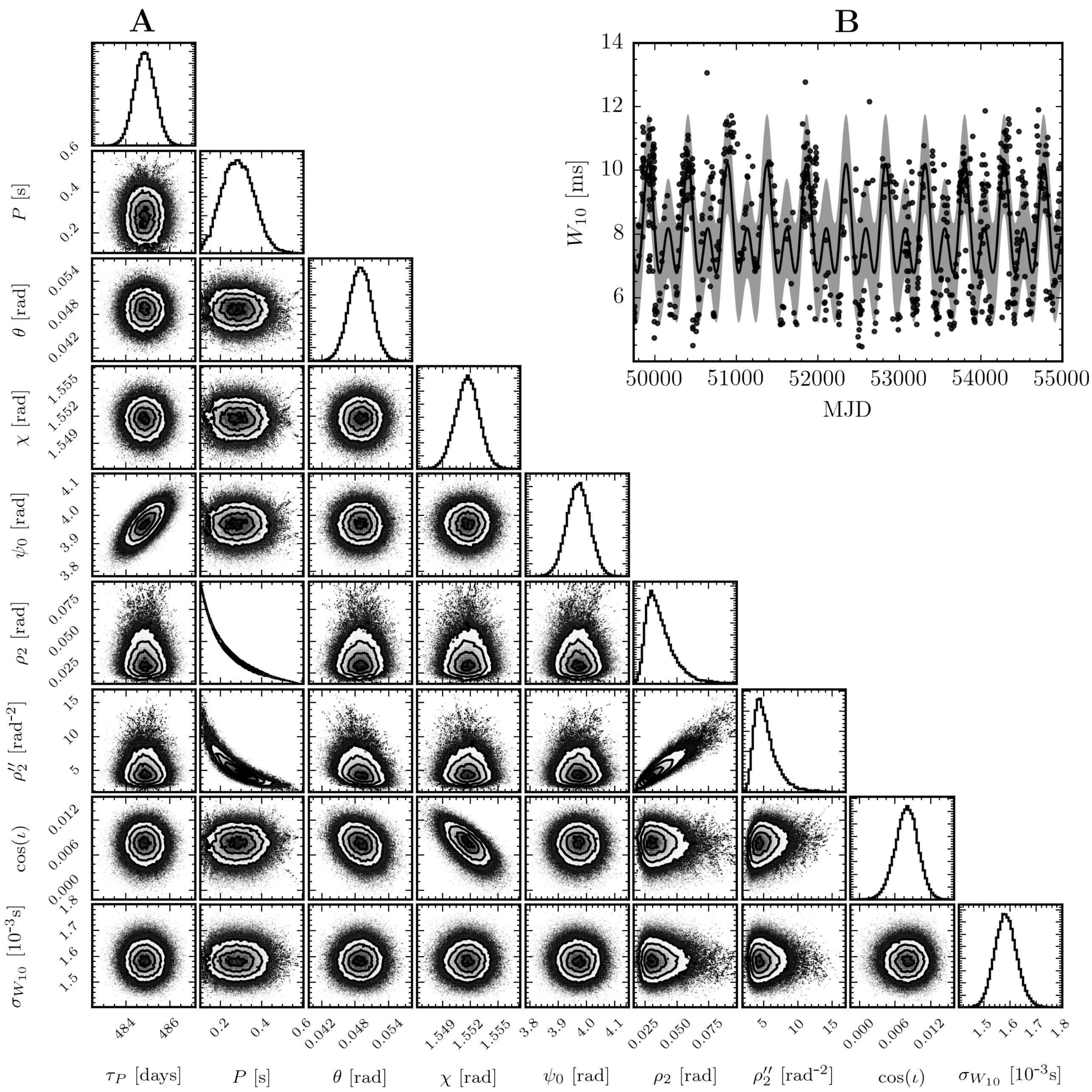

Figure 12. **A**: The estimated marginal posterior probability distribution for the Modified-Gaussian Precession beam-width model parameters. **B**: Checking the fit of the model using the maximum posterior values to the data, see Fig. 3 for a complete description.

modal distributions. In this work we use this method not to help with the efficiency of the simulations [2], but instead so that we can apply the method prescribed by Goggans & Chi (2004) to estimate the evidence as follows.

First we define the inverse temperature $\beta = 1/T$, then let the marginal likelihood as a function of β be

$$Z(\beta) = \int P(\text{data}|\boldsymbol{\theta})^{\beta} P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{36}$$

When $\beta = 1$, this gives exactly the marginal likelihood first defined in Eqn. (3). After some manipulation we see that

$$\frac{1}{Z}\frac{\partial Z}{\partial \beta} = \frac{\int \ln(P(\text{data}|\boldsymbol{\theta}) P(\text{data}|\boldsymbol{\theta})^{\beta} P(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int P(\text{data}|\boldsymbol{\theta})^{\beta} P(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{37}$$

From this, we note that the right-hand-side is an average of

[2] All the posteriors are either unimodal or multimodal with little separation between the modes

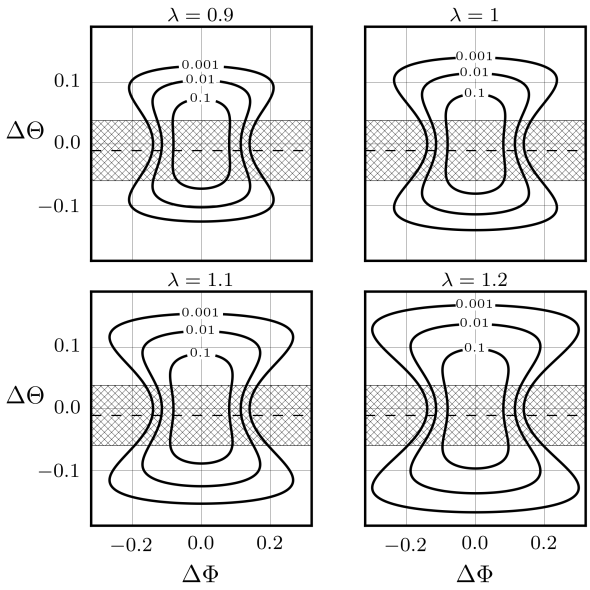

Figure 13. Recreating the beam geometry from the MPE of the Modified-Gaussian Precession beam-width model parameters for four different values of λ. Thick black lines indicate contour lines of the intensity function at fractions of the maximum intensity. The hatched area indicates the region of horizontal cuts (pulses) sampled by the observer: this has a mean, $\chi - \iota$, close to zero (marked by a dashed line) and varies by $\pm\theta$ about this mean.

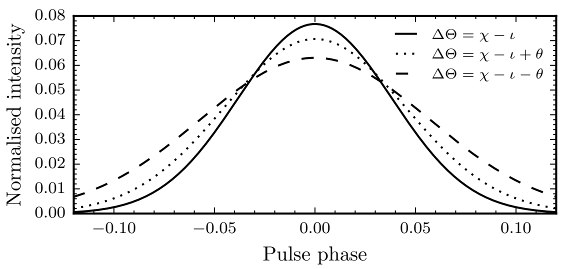

Figure 14. Recreating the pulse profiles for three particular slices through the beam using a fixed value of $\lambda = 1$.

the log-likelihood at β and so

$$\frac{\partial}{\partial \beta}\left(\ln(Z)\right) = \langle \ln(P(\text{data}|\boldsymbol{\theta}))\rangle_{\beta}. \tag{38}$$

Using the likelihoods calculated in the MCMC simulations, we numerically integrate the averaged log-likelihood over β which yields an estimation of the marginal likelihood. To be confident that the estimate is correct, we ensure that we use a sufficient number of temperatures and that they cover the region of interest.

4.2 Results

Applying the thermodynamic integration technique to all the models, we estimate the evidence for each model. Taking the ratio of the evidences gives us the Bayes factor and since we set the ratio of the prior on the models to unity, the Bayes factor is exactly the odds ratio (see Eqn. (2)).

Table 10. Tabulated log-odds-ratios for all models. *By the precession model here we mean the precession with a modified Gaussian beam model as discussed in Sec. 3.3.5.

Model A	Model B	$\log_{10}$(odds-ratio)
Switching	Noise-only	57.0 ± 0.9
Precession*	Noise-only	60.8 ± 0.8
Precession*	Switching	3.8 ± 0.9

We present the $\log_{10}$ odds-ratio between the models in Table 10. A positive value indicates that the data prefers model A over model B. Note that the error here is an estimate of the systematic error due to the choice of β values (see Foreman-Mackey et al. 2013, for details).

This table allows quantitative discrimination amongst the models. The first two rows compare the Switching and Modified-Gaussian Precession models against the Noise-only model with the periodic modulating models being strongly preferred in both cases. Then in the last row we present the log-odds-ratio between the Modified-Gaussian Precession and Switching model which shows that the data prefers the Modified Gaussian model by a factor $10^{3.8}$. Using the interpretation of Jeffreys (1998), the strength of this evidence can be interpreted as "decisive" in favour of this precession model. For completeness, we also mention that the odds-ratio for the non-modified Gaussian model (which failed to fit the data in a physically meaningful way) against the Noise-only model was 3.4 ± 2.4.

5 DISCUSSION

In this work we are using a data set (provided by Lyne et al. (2010)) on the spin-down and beam-width of B1828-11 to compare models for the observed periodic variations. The two concepts under consideration are free precession and magnetospheric switching. In order to be quantitative, we built signal models for the beam-width and spin-down from these conceptual ideas. Using the spin-down data to create proper, physically motivated priors for the beam-width parameters, we then perform a Bayesian model comparison between the models asking "which model does the beam-width data support?". For the models considered here, the data most strongly supports a precession model with a modified Gaussian beam-geometry allowing for both an elliptical beam where the ellipticity has a latitudinal dependence.

To be clear, this does not rule out the switching interpretation since we have not tested an exhaustive set of models—we can only compare between particular models. As an example we could imagine modifying the switching model such that either the switching times, or the magnetospheric states are probabilistic (or a combination of the two). Further we believe there is good grounds to develop models combining the precession and switching interpretation like those discussed in Jones (2012).

The process of fitting the models to the data and performing posterior predictive checks also provides a mechanism to evaluate the models. For both spin-down models the maximum posterior plots with the data (Figs. 5B and 7B) revealed a systematic failure to fit the second (slightly lower) minima. This suggests new ingredients could be introduced to both models to explain this. In Fig. 8 we discovered that the posterior of the spin-period P, as found from fitting a precession model to the spin-down data, has a mode remarkably close to the known value. This provides an additional consistency test for the precession model.

The posteriors for the precession model indicate that B1828-11 is a near-orthogonal rotator and we observe it from close to the equatorial plane. If this is the case, and the two beams of the pulsar are symmetric about the origin, then we expect to see the second beam as an interpulse. Indeed, we discuss further in Appendix B how during the precessional cycle we should expect the intensity of this second beam to dominate at certain phases. Since no such interpulse is reported, either the second beam is weaker, or the beams must have a kink of greater than $4.6°$.

In this work we have developed the framework to evaluate models for the variations observed in B1828-11. This is not intended as an exhaustive review of all models, but rather a discussion on the intricacies that arise such as setting up proper and well-motivated priors. This work lays the groundwork for a more exhaustive test of all available models and can also be extended by improvements to the data sources.

ACKNOWLEDGEMENTS

GA acknowledges financial support from the University of Southampton and the Albert Einstein Institute (Hannover). DIJ acknowledge support from STFC via grant number ST/H002359/1, and also travel support from NewCompStar (a COST-funded Research Networking Programme). We especially thank Will Farr, Danai Antonopoulou, and Ben Stappers for valuable discussions and comments, Foreman-Mackey et al. (2014) for the software used in generating posterior probability distributions, and Lyne et al. (2010) for generously sharing the data for B1828-11.

REFERENCES

Akgün T., Link B., Wasserman I., 2006, MNRAS, 365, 653
Arzamasskiy L., Philippov A., Tchekhovskoy A., 2015, preprint, (arXiv:1504.06626)
Bailes M., Lyne A. G., Shemar S. L., 1993, in Phillips J. A., Thorsett S. E., Kulkarni S. R., eds, Astronomical Society of the Pacific Conference Series Vol. 36, Planets Around Pulsars. pp 19–30
Cordes J., 1993, in Planets around pulsars. pp 43–60, http://adsabs.harvard.edu/full/1993ASPC...36...43C
Cordes J. M., 2013, ApJ, 775, 47
Davis L., Goldstein M., 1970, The Astrophysical Journal, 159
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Foreman-Mackey D., Price-Whelan A., Ryan G., Emily Smith M., Barbary K., Hogg D. W., Brewer B. J., 2014, triangle.py v0.1.1, doi:10.5281/zenodo.11020, http://dx.doi.org/10.5281/zenodo.11020
Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, Bayesian data analysis. CRC press
Goggans P. M., Chi Y., 2004, in AIP Conference Proceedings. pp 59–66
Goodman J., Weare J., 2012, Comm. App. Math. Comp. Sci., 5
Jaynes E. T., 2003, Probability theory: the logic of science. Cambridge university press
Jeffreys H., 1998, The theory of probability. OUP Oxford
Jones D. I., 2012, MNRAS, 420, 2325
Jones D. I., Andersson N., 2001, MNRAS, 324, 811
Landau L. D., Lifshitz E. M., 1969, Mechanics. Pergamon press
Link B., 2003, Physical Review Letters, 91, 101101
Link B., Epstein R. I., 2001, ApJ, 556, 392
Lyne A. G., Manchester R. N., 1988, MNRAS, 234, 477
Lyne A., Hobbs G., Kramer M., Stairs I., Stappers B., 2010, Science, 329, 408
MacKay D. J., 2003, Information theory, inference and learning algorithms. Cambridge university press
Maciesiak K., Gil J., Ribeiro V. A. R. M., 2011, MNRAS, 414, 1314
Manchester R. N., Lyne A. G., 1977, MNRAS, 181, 761
Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, The Astronomical Journal, 129, 1993
Narayan R., Vivekanand M., 1983, A&A, 122, 45
Perera B. B. P., Stappers B. W., Weltevrede P., Lyne A. G., Bassa C. G., 2015, MNRAS, 446, 1380
Radhakrishnan V., Cooke D. J., 1969, Astrophys. Lett., 3, 225
Seymour A. D., Lorimer D. R., 2013, MNRAS, 428, 983
Shaham J., 1977, ApJ, 214, 251
Sivia D., Skilling J., 1996, AMC, 10, 12
Stairs I., Lyne A., Shemar S., 2000, Nature, 406, 484
Swendsen R. H., Wang J.-S., 1986, Phys. Rev. Lett., 57, 2607

APPENDIX A: PROCEDURE FOR MCMC PARAMETER ESTIMATION

The procedure used to simulate the posterior distribution can determine the quality of the estimation. Therefore we will now set out an algorithmic method to ensure the our results are reproducible.

To estimate the posterior given a signal model and prior, we run two MCMC simulations: an initialisation and production. In the following the term *walker* refers to single chains in the MCMC simulation, the Foreman-Mackey et al. (2013) implementation runs a number of these in parallel for each simulation.

- For the initialisation run, we draw samples from the prior distribution to set the initial parameters for each walker. The simulation therefore has the chance to explore the entire parameter space. After a sufficient number of steps, the walkers will converge to the local maxima in the log-likelihood. By visually inspecting the data we determine the nature of the local maxima: in all cases a single maxima dominated such that, given a sufficient length of simulation, we expect all walkers to converge to this maxima. Alternatively we could have found multiple similarly strong maxima, in this case further analysis would be required. This was not found to be the case for any of the models in this analysis.
- For the second step we set the initial state of 100 walkers by uniformly dispersing them in a small range about the maximum-likelihood found in the previous step. The simulation proceeds from this initial state and we divide the resulting samples equally into two: discarding the first half as a so-called 'burn-in', we retain the second half as the production data used to estimate the posterior. The burn-in removes any memory of the artificial initialisation of the walkers at the start of this step.

Having run an MCMC simulation we check that the chains have properly converged (for a discussion on this see Gelman et al. (2013)). The MCMC simulations provide an estimate of the posterior densities for the model parameters. We will also perform 'posterior predictive checks' to ensure the posterior is a suitable fit to the data, i.e. we compare the data to the model prediction when the model parameters are set to the values corresponding to the peaks of the posterior probability distributions.

APPENDIX B: IMPLICATIONS FOR THE UNOBSERVED BEAM

The precession model developed here assumes the observer only ever sees one pole of the beam-axis, but in the canonical model we often imagine there is also emission from an opposite magnetic pole. In several pulsars this can be seen as an *interpulse* 180° out of phase from the main pulse (Lyne & Manchester 1988; Maciesiak et al. 2011); these pulsars are generally found to be close to orthogonal rotators[3]. No such interpulse is reported for B1828-11, we will now discuss the implications of this given the precession interpretation.

Let us imagine a scenario where the observer is in the northern (magnetic) hemisphere and label the beam protruding into their hemisphere (which they will see with the greater intensity due to the smaller angular separation) at the start of the thought experiment as the north pole. Then the north and south pole make angles Θ and $\pi - \Theta$ respectively with the fixed angular momentum vector. Now we see that if, during the course of the precessional cycle, $\Theta > \pi/2$ then the south pole will protrude into the northern hemisphere and the north pole into the southern hemisphere. Provided both poles are identical, but regardless of the details of the beam-geometry, at this time we must expect the observer to see the south pole at greater intensity than the north pole. An example of this is shown in Fig. B1, but note that the observer will see the greatest intensity from the south-pole half a rotation after this instance.

Our posterior distributions inform us that, if the precession interpretation is correct, we are in exactly this situation: Θ ranges from 85.8° to 92.3° over a precessional cycle[4] so we should see the interpulse.

This is readily explained if the south pole is substantially weaker in intensity, or by the one-pole interpretation Manchester & Lyne (1977). Alternatively, it could be that the two beams are not diametrically opposed but are latitudinally 'kinked'. In the later case we can put a lower bound on the kink angle by requiring that the polar angle of the south pole is always greater than that of the north (see Fig. B1). From our MPE this gives a lower bound of 4.6° for the polar kink angle. This latitudinal kink can be compared with the longitudinal kink of interpulses observed in other pulsars: often these are not found at exactly 180°, but can deviate by 10's of degrees (see the separation of interpulses for double-pole interpulses in Table 1. of Maciesiak et al. (2011)).

[3] The use of interpulse here strictly refers to seeing the opposite beam of the pulsar, and not cases where the pulsar is almost aligned and interpulses are thought to come from the same beam
[4] Numbers generated from the maximum posterior estimates of χ and θ using the spin-down data. The estimated error for both values is ±0.08°

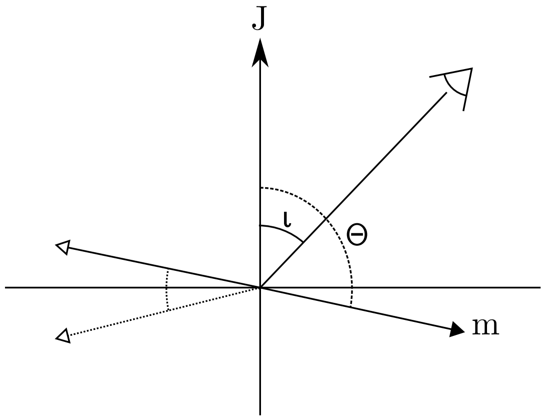

Figure B1. Extension to Fig. 9 adding in the south pole (demarked by a white triangle and solid line) for an instance in which the north pole is in the southern hemisphere. The dotted line with a white arrowhead is the 'kinked' beam. For clarity, the north and south poles are shown at times that differ by half a rotation period.

APPENDIX C: ASYMMETRY OF THE MINIMA IN THE SPIN-DOWN DATA: IMPLICATIONS FOR PRECESSION

From the observational data for the spin-down in Fig. 1, we can see that the two minima are not of equal depth (ignoring the effect of the secular $\ddot{\nu}$). In this section we will briefly discuss the implications of this for precession.

In Eqn. (14) we neglected the geometric modulation of the spin-down since both Jones & Andersson (2001) and Link & Epstein (2001) demonstrated it can be expected to be negligible for B1828-11. This geometric term can be found explicitly in the second term Eqn. (20) from Link & Epstein (2001), we write it now in the following form

$$\Delta\dot{\nu}_{\rm geo}(t) = \frac{2\pi}{\tau_{\rm P}^2}\left(\theta\cot\chi\cos\psi(t) + \theta^2(1+2\cos^2\chi)\sin(2\psi(t))\right). \tag{C1}$$

Let us also extract from Eqn. (14) the electromagnetic modulation

$$\Delta\dot{\nu}_{\rm em}(t) = \frac{1}{2\pi\tau_{\rm Age}P}\left(2\theta\cot\chi\sin\psi(t) - \frac{\theta^2}{2}\cos(2\psi(t))\right). \tag{C2}$$

We want now to consider only the form of these equations, so we shall write them in a generalised way:

$$\Delta\dot{\nu}_{\rm geo}(t) = A_{\rm geo}\cos\psi(t) + B_{\rm geo}\sin(2\psi(t)), \tag{C3}$$
$$\Delta\dot{\nu}_{\rm em}(t) = A_{\rm em}\sin\psi(t) - B_{\rm em}\cos(2\psi(t)), \tag{C4}$$

and use labels A_i and B_i, where i = geo or i = em. Each of these expression can be phase-shifted by adding ψ_0. It may be thought that under such a transformation, the two expression can be shown to be equivalent, but this is not the case and they are in fact very different in character.

Let us consider when $A_i \sim B_i$ such that both equations demonstrate doubly-periodic behaviour and will have four turning points (in the limit $B_i \ll A_i$ or $B_i \gg A_i$ both expression are equivalent up to a phase-shift). Then solving for the minima and maxima algebraically, we find that for $\Delta\dot{\nu}_{\rm em}$ two of the extrema will always take the same y-value.

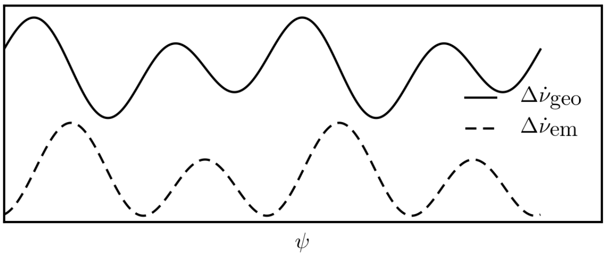

Figure C1. A demonstration of the characteristic form of Eqn. (C3) and (C4) over two periods. The offset in the y-axis is intended only to help guide the eye, both modulate the spin-down about 0.

This can be understood to be a consequence of the symmetry

$$\Delta\dot{\nu}_{\rm em}(\pi - \psi) = \Delta\dot{\nu}_{\rm em}(\psi). \tag{C5}$$

No such symmetry is found for $\Delta\dot{\nu}_{\rm geo}$.

To provide insight into this, we plot in Fig. C1 both equations for particular instances of A_i and B_i over a two periods. This demonstrates the four turning points for both equations, and the repeated y-value for (in this instance) the two minima of $\Delta\dot{\nu}_{\rm em}$.

The spin-down data by-eye clearly indicates that it will be best fit by a modulation which has four unequal turning points. As we have demonstrated, Eqn. (14), which uses only the electromagnetic modulation $\Delta\dot{\nu}_{\rm em}$ is subject to a constraint on two of its extrema being equal; therefore it will never properly fit the data since it must repeatedly fail to fit one of the extrema.

The geometric term on the other hand does not suffer this constraint. If we attempt to refit the data simply adding on the geometric modulation (which was previously neglected) we find that, for B1828-11, the geometric modulation as given in Eqn. (C1) is negligible confirming the results of Jones & Andersson (2001) and Link & Epstein (2001). This does however pose a question: is there some physics missing which may suppress the electromagnetic amplification, or is their even a different modulation altogether?

This paper has been typeset from a TEX/LATEX file prepared by the author.